

## Appendix 2

### R code `val.prob.ci.2`

The function `val.prob.ci.2` is an adaptation of `val.prob` from Frank Harrell's `rms` package, <https://cran.r-project.org/web/packages/rms/rms.pdf>. The key feature of `val.prob.ci.2` is the generation of logistic and flexible calibration curves and related statistics. It is used as follows:

```
val.prob.ci.2 <- function(p, y, logit, group, weights=rep(1,
  length(y)), normwt=F, pl=T, smooth=c("loess","rcs"),
  CL.smooth=F, CL.BT=F, nr.knots=5, logistic.cal=T,
  xlab="Predicted probability", ylab="Observed proportion",
  xlim=c(-0.02, 1), ylim=c(-0.15,1), m, g, cuts, emax.lim=c(0, 1),
  legendloc=c(0.50 , 0.27), statloc=c(0, .85), dostats=T,
  roundstats=2, riskdist="predicted", cex=0.75, cex.leg=0.75,
  mkh=0.02, connect.group=F, connect.smooth=T, g.group=4,
  evaluate=100, nmin=0, d0lab="0", d1lab="1", cex.d01=0.7,
  dist.label=0.04, line.bins=-.05, dist.label2=.03, cutoff, las=1,
  length.seg=1, ...)
```

New options, or options that are adapted from the original `val.prob` are (default values underlined):

`smooth`        "loess" generates a flexible calibration curve based on loess, "rcs" generates a calibration curves based on restricted cubic splines, F or FALSE suppresses the flexible curve. We recommend to use loess unless N is large, for example  $N > 5000$ .

`CL.smooth` "fill" shows pointwise 95% confidence limits for the flexible calibration curve with a gray area between the lower and upper limits, `T` or `TRUE` shows pointwise 95% confidence limits for the flexible calibration curve with dashed lines, `F` or `FALSE` suppresses the confidence limits.

`CL.BT` `T` or `TRUE` uses confidence limits based on 2000 bootstrap samples, `F` or `FALSE` uses closed form confidence limits.

`nr.knots` specifies the number of knots for rcs-based calibration curve. The default as well as the highest allowed value is 5. In case the specified number of knots leads to estimation problems, then the number of knots is automatically reduced to the closest value without estimation problems.

`dostats` specifies whether and which performance measures are shown in the figure. `T` or `TRUE` shows the "abc" of model performance (Steyerberg et al, Rev Esp Cardiol 2011): calibration intercept, calibration slope, and c statistic. `F` or `FALSE` suppresses the presentation of statistics in the figure. A `c()` list of specific stats shows the specified stats. The key stats which are also mentioned in this paper are "C (ROC)" for the c statistic, "Intercept" for the calibration intercept, "Slope" for the calibration slope, and "ECI" for the estimated calibration index (Van Hoorde et al, J Biomed Inform 2015). The full list of possible statistics is taken from `val.prob` (Harrell, 2001) and augmented with the estimated calibration index: "Dxy", "C (ROC)", "R2", "D", "D:Chi-sq", "D:p", "U", "U:Chi-sq", "U:p", "Q", "Brier", "Intercept", "Slope", "Emax", "Brier scaled", "Eavg", "ECI". These statistics are always shown in the log window.

`roundstats` specifies the number of decimals to which statistics are rounded when shown in the figure. Default is 2.

`cex, cex.leg` controls the font size of the statistics (`cex`) or plot legend (`cex.leg`). Default is 0.75.

`d0lab, d1lab` controls the labels for events and non-events (i.e. outcome  $y$ ) for the histograms. Defaults are `d1lab="1"` for events and `d0lab="0"` for non-events.

`cex.d01` controls the size of the labels for events and non-events. Default is 0.7.

`dist.label1` controls the horizontal position of the labels for events and non-events. Default is 0.04.

`dist.label2` controls the vertical distance between the labels for events and non-events. Default is 0.03.

`line.bins` controls the horizontal (y-axis) position of the histograms. Default is -0.05.

`cutoff` puts an arrow at the specified risk cut-off(s). Default is none.

`las` controls whether y-axis values are shown horizontally (1) or vertically (0).

`length.seq` controls the length of the histogram lines. Default is 1.

Further options relevant to this paper:

`logistic.cal` T or TRUE plots the logistic calibration curve, F or FALSE suppresses this curve.

`cuts` provides a list `c()` of actual cutpoints for categorization, in case calibration should be shown for categorizations of predicted risk.

m provides average size of categories, in case calibration should be shown for categorizations of predicted risk.

g provides number of equally large (quantile) categories, in case calibration should be shown for categorizations of predicted risk. E.g. g=10 groups predicted risk using deciles as is commonly done.

group this provides a stratification variable for which to stratify the calibration analysis. This can be used to get average predicted risks and observed proportions for subgroups of patients based on the stratification variable. T or TRUE performs this analysis for the whole dataset. A plot with loess calibration curves per subgroup can be obtained with

```
plot(val.prob.ci.2(p, y, group=z)).
```

g.group If the stratification variable is continuous, g.group defines the number of quantile groups (default is quartiles, i.e. 4).

riskdist indicates whether histograms are based on predicted risk ("predicted") or on calibrated risk ("calibrated").